

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Experimental Self-calibration from Four Views

P.D. 13.09.1995 P: 307-312 ⑤

Reyes Enciso and Thierry Viéville

Projet Robotvis, INRIA Sophia-Antipolis, B.P. 93
F-06902 Sophia-Antipolis, France.

Abstract. The main goal of self-calibration [2, 3, 8] is to compute the intrinsic and extrinsic parameters of a camera without using a known pattern. In this paper we focus on the calibration of a binocular head-eye system from four views. The only information provided to the algorithm is the fundamental matrices [3] and the point correspondences between the 4 views. We exploit the information of the cross-correspondences to improve the Euclidean reconstruction.

1 Introduction

In this paper we address the problem of computing the intrinsic and extrinsic calibration parameters, in a binocular image sequence, given a set of point correspondences. Most authors [2, 5, 8] studied the case of point correspondences, but have restricted their approach to the case where the intrinsic parameters of the camera are constant, while only 2 or 3 views have been taken into account, or studied the monocular case for the long sequences as in [7].

The generalization to the case where intrinsic parameters are non-constant has already been addressed. But usually the analysis is restricted to the recover of affine or projective structure of the scene.

This paper extends these previous works to the case of non-constant intrinsic parameters and non-constant relative positioning of the cameras of the stereoscopic system. In particular, in the case of active vision, the extrinsic and intrinsic parameters of the visual sensor are modified dynamically. For instance, when tuning the zoom and focus of a lens, these parameters are modified and must be considered as dynamic parameters. It is thus necessary to attempt to determine dynamic calibration parameters by a simple observation of an unknown stationary scene, when performing a rigid motion.

2 The Camera Model

We use the well-known pinhole camera model (see Fig. 1) [3]. We assume that there will be a perfect perspective projection with center C (the *optical center*) at a distance f (*focal distance*) from the *retinal plane* \mathcal{R} . The plane containing the optical center is called the *focal plane*.

Three main coordinate frames are defined: the **world coordinate frame**, the **camera frame** defined by its origin C and the axis (X_c, Y_c, Z_c) , and the

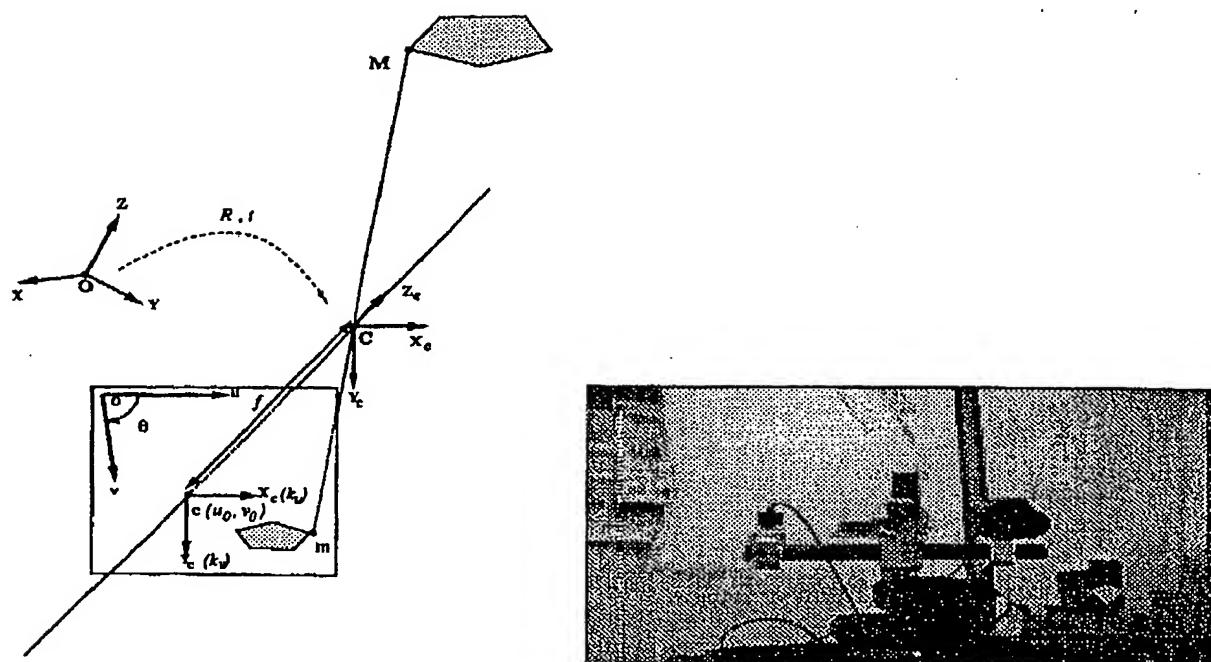


Fig. 1. The pinhole model and our binocular head.

normalized frame at focal distance equal to 1, which origin is called the *principal point* c , and axis (X_c, Y_c) .

Using projective coordinates, the 3D world point $M = [X, Y, Z]$ and its retinal normalized image $m = [u, v]$, are related by

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \underbrace{A \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} D}_{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad (1)$$

where A is the matrix of the *intrinsic parameters* and D the matrix of the *extrinsic parameters*:

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}, D = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}.$$

We have chosen to represent the 3x1 translation vector t in spherical coordinates (t is defined up to a scale factor and chosen to be unitary), the rotation matrix R with the Rodrigues formulation [8] and A depending only on four parameters α_u, α_v, u_0 and v_0 . Other authors use a fifth parameter θ , which measures the non-orthogonality of the pixels, but its value is always less than noise [1].

3 The Fundamental Matrix.

As already demonstrated by Faugeras in [2] the equation of Longuet-Higgins relating the fundamental matrix F and the point correspondences q and q'

between two views can be written:

$$\mathbf{q}'^t \mathbf{F} \mathbf{q} = 0, \quad (2)$$

with:

$$\mathbf{F} = \mathbf{A}'^{-1t} \mathbf{E} \mathbf{A}^{-1}, \quad (3)$$

where $\mathbf{E} = \mathbf{T}\mathbf{R}$ is the essential matrix. \mathbf{T} is a skew symmetrical matrix defined by the translation vector \mathbf{t} such that $\mathbf{T}\mathbf{x} = \mathbf{t} \wedge \mathbf{x}$ for all 3D vector \mathbf{x} (\wedge denotes the cross-product). Of course, $\mathbf{E} = \mathbf{A}^t \mathbf{F} \mathbf{A}$ when both cameras are identical. For any pair of views we are trying to compute: 3 parameters for the rotation, 2 for the translation (defined up to a scale factor), and 4 intrinsic parameters. Since the essential 3x3 matrix $\mathbf{E} = \mathbf{T}\mathbf{R}$ is of rank two, and because \mathbf{A} and \mathbf{A}' are invertible, \mathbf{F} is of rank two. Then, we can compute 7 parameters from each fundamental matrix: the motion between the two views and 2 intrinsic parameters. The principal point (u_0, v_0) is fixed, and only the factors α_u and α_v need to be estimated.

4 Calibrating the Binocular Head-Eye System

We use a binocular head-eye system (Fig. 1). A description of this active visual system can be found in [6]. To calibrate the head-eye we need to know the intrinsic parameters of each camera at any position (in this study \mathbf{A}_i with $i = 1..4$), the displacement of each camera (right \mathbf{D}_r and left \mathbf{D}_l), the displacement between both cameras (stereo displacements \mathbf{D} and \mathbf{D}'), and finally the cross-displacements ($\mathbf{D}^o, \mathbf{D}^*$). From Fig. 2 we can easily write: $\mathbf{D}'\mathbf{D}_l = \mathbf{D}_r\mathbf{D}$, $\mathbf{D}^o = \mathbf{D}_r\mathbf{D}$ and $\mathbf{D}_r = \mathbf{D}'\mathbf{D}^*$. This means that we can compute all rotation matrices and translation vectors from, for instance, $\mathbf{R}_r, \mathbf{t}_r, \mathbf{R}_l, \mathbf{t}_l, \mathbf{R}'$ and \mathbf{t}' .

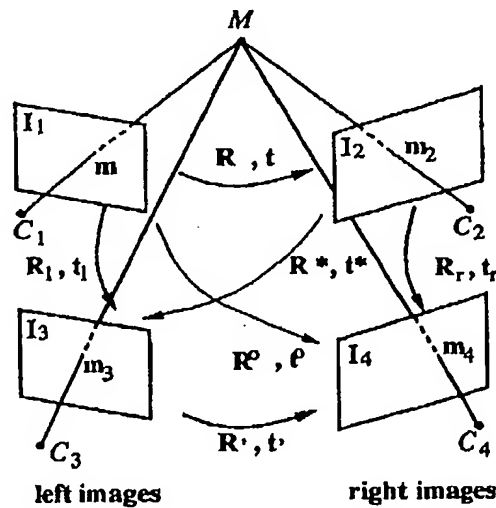


Fig. 2. Our stereoscopic system

4.1 The Non-Linear Minimization Algorithm

The **input** is the six fundamental matrices relating the four views (or at least the first four): $F_{12}, F_{13}, F_{24}, F_{34}, F_{14}, F_{23}$, and the correspondences. We denote m_{ij}^{ik} as the j -th correspondence in image i , and between the images i and k . As explained in Sect. 3, we obtain from each F an estimation of R and t once the intrinsic parameters are known. We compute the **output** intrinsic and extrinsic parameters replacing in the next criterion each F by the equation 3, and minimizing:

$$\min \left[\sum_{i=1}^p ((m_{2i}^{21})^t F_{12} m_{1i}^{12})^2 + \sum_{j=1}^q ((m_{3j}^{31})^t F_{13} m_{1j}^{13})^2 + \sum_{k=1}^r ((m_{4k}^{42})^t F_{24} m_{2k}^{24})^2 + \right. \\ \left. \sum_{l=1}^s ((m_{4l}^{43})^t F_{34} m_{3l}^{34})^2 + \sum_{m=1}^t ((m_{4m}^{41})^t F_{14} m_{1m}^{14})^2 + \sum_{n=1}^u ((m_{3n}^{32})^t F_{23} m_{2n}^{23})^2 \right]$$

4.2 The Parameters Computed depending on the Model Used.

The **Model 0**, or the simplest model, assumes that $D = D'$ and the intrinsic parameters are constant. This is the case when the vergence of the right camera, the zoom and the focus are not changed. The principal point is fixed to the center of the image: (255, 255) pixels. To reduce the number of intrinsic parameters (**Model 1**), we know [1] that the quotient $C_0 = \frac{\alpha_u}{\alpha_v} = 0.7$ is constant, so we can write the intrinsic parameter matrix for each camera position i :

$$A = \begin{pmatrix} C_0 \cdot \alpha_v & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using the **Model 2**, we do not a priori set any parameter as being constant. If N is the number of views [7], we have $11N-15$ independent parameters, which is in our case $11 \cdot 4 - 15 = 29$. In the table 1 we have represented the different models. We denote $i=1, \dots, N$ as the index of the view.

Table 1. The Euclidean parameters computed depending on the model

EUCLIDEAN PARAMETERS					
Model	intrin	rotation	trans	TOTAL	N = 4
Model 0 u_0, v_0 fixed A_i fixed, $D = D'$	4	$\frac{3N}{2}$	$\frac{3N}{2} - 1$	$3N+3$	15
Model 1 $\frac{\alpha_u}{\alpha_v} = 0.7$	$3N$	$3(N-1)$	$3(N-1)-1$	$9N-7$	29
Model 2	$4N$	$3(N-1)$	$3(N-1)-1$	$10N-7$	33

4.3 Euclidean Reconstruction Results Using Real Images

Two minimization routines have been used: *e04fcf()* of NAG and *frprmn()* of Numerical Recipes (NR). The parameters are initialized to their values by default: (u_0, v_0) are initialized to (255, 255) and (α_u, α_v) are set to (800, 800). The extrinsic parameters are computed by developing (3), once the intrinsic parameters are set. In the Fig. 3, 4 we show different views of the 3D Euclidean reconstructed segments of the scene.

SEQUENCE 1: Intrinsic and stereo parameters are constant. Between the two pairs only the elevation of the stereo frame changes. The correspondences are extracted automatically using the grid.

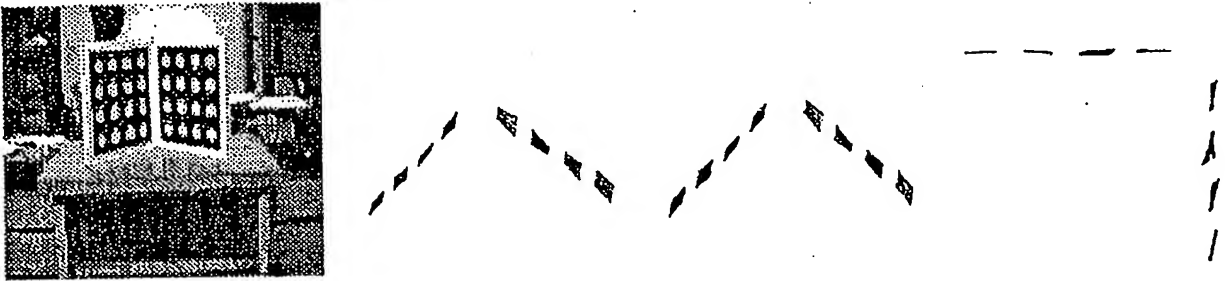


Fig. 3. From the left to the right: *Camera View* of the superimposed 3D Euclidean reconstruction using model 1 and the library NAG when the parameters are initialized to their default values, *Top View* of 3D Euclidean reconstruction result of our algorithm when the parameters are initialized to their default values, initialized by the result of "hard calibration" [4], and the result of the "hard calibration" method, respectively.

SEQUENCE 2: Intrinsic and stereo parameters change between the two views. We focused, zoomed and changed the vergence's value. The reconstruction using model 0 and 1 is insatisfactory. The parameters are initialized to their default values. The correspondences are extracted by hand using the sub-pixel accuracy.

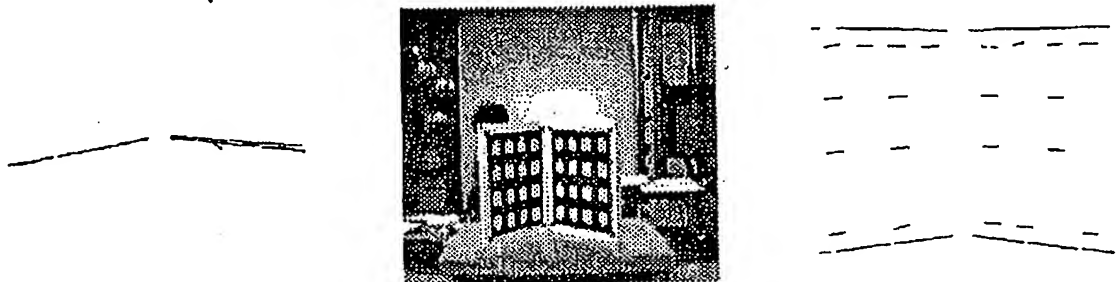


Fig. 4. *Top, Camera and Front View* of the 3D Euclidean reconstruction using the model 2 with the library NR (the results with NAG are similar).

5 Conclusions

We want to point out that although we are using a calibration grid for a convenient feature detection, only sequence 1 uses the 3D model of the grid to increase the number of correspondences and to compare the results with a "hard calibration" method. The other sequence use 50 correspondences. In every case we are using a self-calibration method, that is to say, without using the 3D model of a known pattern in order to compute the perspective projection matrix. The segments are reconstructed automatically.

The most relevant conclusions are: a) The simplified model 0 is not suitable at all when extrinsic or intrinsic parameters change. To set the principal point to the center of the image when using a camera with a zoom is not realistic at all. b) The model 1 is convenient when we do not use the zoom, then the intrinsic parameters $\alpha_u = 0.7 * \alpha_v$. It works better than the model 2 because there are 29 unknowns (the same number of independent parameters).

The gradient conjugate method's results are worse than Newton's modified method results for real images.

In the case the stereo displacement does not change between the two frames, if we initialized the first frame's parameters with the calibration hard method, then the results are very good. In the case in which the stereo parameters change, the algorithm minimizes the epipolar distance but increases the reconstruction error.

References

1. R. Enciso, T. Viéville, and O. Faugeras. Approximation du changement de focale et de mise au point par une transformation affine à trois paramètres. *Traitement du Signal*, 11(5), 1994.
2. O. Faugeras, Q.-T. Luong, and S. Maybank. Camera self-calibration: theory and experiments. In *2nd ECCV*, pages 321–334, Santa-Margherita, Italy, 1992.
3. T. Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement*. PhD thesis, Université de Paris-Sud, Orsay, 1992. PhD thesis.
4. L. Robert. *Perception Stéréoscopique de Courbes et de Surfaces Tridimensionnelles, Application à la Robotique Mobile*. PhD thesis, Ecole Polytechnique, Palaiseau. France, 1992. PhD thesis.
5. T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.
6. T. Viéville, E. Clergue, R. Enciso, and H. Mathieu. Experimenting 3d vision on a robotic head. In *The 12th Int. Conf. on Pattern Recognition*, pages 739–743, 1994.
7. T. Viéville, Q. Luong, and O. Faugeras. Motion of points and lines in the uncalibrated case. *International Journal of Computer Vision*, 1994. To appear.
8. Z. Zhang, Q. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. Technical Report 2079, INRIA, 1993.